Making Visible
Illustration Through Identification, Categorization, and Metaphor
Stuart Medley

Introduction

When it comes to communicating visually, there are essentially two means of making pictures, the photographic (which also includes film and video) and the chirographic (including any hand-made means of depicting), and degrees of hybridity between. The first category has relatively narrow bandwidth. The second is extremely broad and is covered by what is referred to in this book as illustration. The chirographic may comprise pictures ranging from caricature to scientific illustration and infographics, from technical drawings and diagrams to the depictions in children’s books and comics. Illustration needs an art theory that places it in the context of visual communication. Illustration is different to art in its intentionality, but even celebrated practitioners struggle to fully articulate its role and effectiveness in communicating an intended message (Medley 2013). It generally attaches to a written text or otherwise purposely imparts information or invokes in its beholder a particular esthetic response. At the same time, perhaps because certain illustrative modes share esthetics with visual art, the visual communication tasks to which illustration is put may be assumed to be less practical and more emotional than those deliberate communication requires. Photography, with its capacity to “capture the real thing” with the visual rhetoric of objectivity, seemed a more logical choice. But if one wants to be logical about how and why visual communication works, then some examination of how humans see and understand what they see is necessary. Interestingly, in spite of the seemingly pragmatic nature of photography, the psychology of seeing is better reflected in the many modes of illustration. The enormously wide array of depictive modes illustrators can adopt, most of which depart in some way from visible reality, paradoxically, more closely parallels the various ways that the mind deals with the images presented to it by the eyes.

Photography is very good at capturing light, color, and detail from the real world in ways that are less mediated by the human hand. It is a visual communication mode
that helps its audience identify particular people and places, and is effective in conveying a visceral sense of “being there.” However, as I will explain, identification is only a small aspect of what humans do with the vision supplied by the eyes to the brain. Categorization, the sorting of objects into broad classifications, is a visual task that the brain is more often dealing with, and one which requires substantially less detail than that found in the real world or recorded through photography.

In making meaning of what is seen, the beholder connects vision to concepts. Identification assumes there is an objective reality “out there” to be got at. However, depending on where the beholder stands in nature, some things are more identifiable than others. Distance and lighting conditions can both reduce the kind of detail associated with visual realism, even “in the wild.” In addition, the eye can never see exactly the same thing twice. The mind must, and does, have ways of rounding out these continuously varying impressions. Illustration allows for these more generalized views of things to be presented to the beholder rather than the sliver of reality captured in the snapshot. This chapter, in part, reviews literature, particularly from perceptual and cognitive science, as it applies to illustration, in order to explain how illustrations can help the beholder with the visual tasks of both identification and categorization.

In the second part of the chapter I will explore how illustration need not be a means of recording aspects of the visible world at all but may also make visible those aspects of life that are not normally apparent to the eye. In life, the visibly real, though it requires much of our mental processing powers, and though sight is often prioritized over the other senses, is not the only reality. Most sensible people would agree that economic collapse, for example, is a reality. Like the wind, however, it cannot be directly photographed. One might photograph its effects. However, the wind can be drawn in an illustration. Comics makers, for example, have conventions to hand for depicting the wind using basic visual schemas. The wind can also be personified as it has been on ancient maps. Similarly, “governments propping up corporations deemed too big to fail” cannot be photographed, but as with the words “propping up” one doesn’t have to look too far inside oneself for a mental sketch to depict financial rescue measures. Making invisible aspects of life visible, based on schemas and metaphor, is one of the great strengths of illustration in addressing visual communication tasks.

**Visual Tasks**

To begin with the visual task of identification, I want to ponder how it is we can understand any image that is not realistic. Throughout human evolution our ancestors have been looking at the world and making sense of it through all its detail. One might reasonably expect therefore that a picture that can reproduce that detailed world faithfully would be ideal for all pictorial communication. The relatively recent invention of photography, since it allows the faithful recording of light reflecting from objects in the real world, would seem to be the best way to make these pictures. Certainly the invention of photography was greeted with a sense that here at last was a way of sharing a view of the world (Green-Lewis 1996, p. 26). However, depending on the visual communication task required, instructional design, for example, or recognition and memory tasks, a photograph is often outperformed by other kinds of pictures. Or as Francis Dwyer, the visual literacy advocate, observed, the use of specific types of visual illustrations to facilitate specific types of objectives significantly
improves achievement (Dwyer 1978, pp. 96–97). These other visual communications tasks, and the strengths of illustration modes to carry them out, will be discussed below. For now I will keep the focus on the faithful representation of the visible world.

Identification

High-fidelity reproductions of the visible world are effective where the visual communication task is to help the beholder identify something specific. For example, environmental portraiture is a mode of photography which records a specific person in their typical, identifiable habitat. From a socio-psychological perspective, identification of other humans is high on the list of visual tasks: “Face recognition represents a potent drive to processes underlying natural selection, since it underpins appropriate interaction with the species most central to our survival, namely other humans” (Wallis 2013). While merely broad categorization of animals, foods, tools, and other objects suffices for survival, correct within-category discrimination is necessary for a reliable face recognition system; the relevant question is not “what is that?” (a face), but rather “who is that?” (which particular face). From the evolutionary perspective, faces may merit neural resources beyond those dedicated to other object classes (Kanwisher et al. 1997; Öhman and Mineka 2001; Tsao and Livingstone 2008).

Much of the psychology of face recognition is still debated. Newport et al. (2016) explain: “Two rival theories of how humans recognize faces exist: (i) recognition is innate, relying on specialized neocortical circuitry, and (ii) recognition is a learned expertise, relying on general object recognition pathways.” However, that same study suggests that animals without the brain module hitherto ascribed the function of face recognition (the neocortex) can be trained to recognize human faces. Wallis (2013), on the other hand, argues for a unified model of face and object recognition: “many of the known effects are actually a symptom of expertise rather than something immutably unique to faces” (Wallis 2013, p. 2). As people gain experience of the world they develop concepts regarding what is visually normal within particular classes of objects. For humans the class of objects in which most people develop the greatest level of discrimination is the face. After a time each person, depending on their experience of others, begins to develop an understanding of a mental norm or range of typical possibilities for faces.

As far as representation and illustration are concerned, it is important to focus less on the specific mental mechanisms that allow the brain to identify individuals and more on the fact that identification is a relatively narrow, albeit important, function of the visual system compared with the broader need to correctly classify other things.

If identification is the visual communication task at hand, photography certainly provides a powerful and efficient way of delivering the information required, but there are at least two caveats to this proposition. One is that even in this task photography can be bettered by one of the chirographic illustration modes, caricature. The other caveat is the problem of specificity that photography always supplies. I will deal with the first caveat first. Susan Brennan defines caricature as:

a graphical coding of facial features that seeks paradoxically to be more like a face than the face itself. It […] amplifies perceptually significant information while reducing less relevant details. The resulting distortion satisfies the beholder’s mental model of what is unique about a particular face. (Brennan 1985, p. 170)
Dror et al. (2008) explain that there are advantages to caricaturing a face including better recall and more efficient identification by the beholder. As the authors explain: “performance is better when exaggerated stimuli are presented rather than a faithful image. This can be understood with respect to a theoretical framework in which caricaturing maximises the distinctiveness and thus minimises any perceptual or representational confusion.” Similarly, Frowd et al. (2015) explore the advantages of caricature in the context of identification for criminal prosecution.

It stands to reason that in order for an illustrated caricature to be understood by an audience, its audience must appreciate what the picture has captured in reference to the personality portrayed. The creator has exaggerated the differences, but the beholder must recognize the exaggerated depiction as representing the specific example. Human minds in general, not just those of cartoonists, must therefore have the ability to caricature images. As with identification of faces generally, there are also two schools of thought on how the mind caricatures faces. The first is termed norm-based caricature (NBC) and the second, exemplar only, absolute coding (ABC). The chief difference in exploring the caricature effect is that NBC assumes that each mind has an experience of faces and stores a norm in memory made from averaging out all the faces encountered. A caricaturist, and simply the mind of the beholder, can exaggerate the differences between a unique face and this stored norm. Where a new face differs from the norm, the mind appears to memorize these differences in a form exaggerated beyond their actual appearance. For instance, a person may seem unusual as a result of having eyes very far apart. For a mind to reach this conclusion it must be comparing the eyes of this newly met face with something. That something, according to NBC, is a stored norm for faces. The brain will exaggerate this difference further still by moving the eyes further apart in the stored memory of that person.

In ABC a face is unique because it lies outside of the cluster of most faces in terms of its configuration of features. To make a caricature following this logic, a conceptual vector derives from the center of the cluster of greatest “exemplar density” to the unique face. The caricature effectively moves the unique features further away still from the cluster, continuing on in the direction of the vector. Theorists themselves acknowledge the difficulty of comparing these models: “What makes the two models so difficult to distinguish empirically is the high correlation between exemplar density and distance from the norm” (Byatt and Rhodes 1998, p. 2458). Theorists are agreed, however, that the caricature effect is a function of the mind. In other words, each person is a caricaturist, if only in their memory; caricature is a universal human faculty, not just a practiced mode of a particular kind of illustrator.

That the mind can better identify a caricatured illustration than a photograph of the real thing raises for psychologists a possibility that caricatures better match the memory representation than an undistorted image (Rhodes 1997; Ward 2015). This ability extends beyond face recognition. Any objects that may be discriminated by difference from a norm or difference from the cluster of “exemplar density” may be caricatured. The criterion being that such a group of objects has a norm or cluster of “exemplar density.” I have used caricatures of typefaces to explain non-face caricature to illustration students (Medley and Mutard, 2017) using Helvetica as the norm font by which to caricature other sans serif typefaces; Dror et al. (2008) have used caricatures of aeroplanes to aid learning. It may take at least some expertise in the particular subject to recognize the caricature and to recognize that a specific example has been caricatured, but in a chicken-and-egg-like situation, Dror et al. (2008) also demonstrated that expertise in a new category of objects can be accelerated through training with caricatures of those objects.
Ramachandran and Hirstein (1999) proposed that this exaggerative approach to picture making evident in caricature illustration can be applied as a general principle for making art: “The purpose of art, surely, is not merely to depict or represent reality – for that can be accomplished very easily with a camera – but to enhance, transcend, or indeed even to distort reality.” They have sought the evolutionary advantage to such representations. It seems that the response to exaggerated stimuli is not only innate in humans but this peak shift effect may have arrived at a much earlier evolutionary branch. There are examples in the natural world of animals responding to artificial, and grossly exaggerated, visual stimuli. In ornithology, Nikolaas Tinbergen’s ethological experiments demonstrate a wide range of visual exaggerations which appear to trigger strong responses in their avian beholders, for example, the oystercatcher’s preference for brooding on eggs that are much too large to be its own (1948). Like many illustrations, these stimuli are so exaggerated that they are literally impossible to see in the real world, and yet these seem to appeal to some image in the mind of the beholder (Gardner and Gardner 2013). It can be said then that sentient creatures have in their minds images which do not exist or, in some cases, cannot exist in the physical world. The special license that illustrators and artists more generally possess is to put down on paper, canvas, or screen these “what if” images.

If photographs can be challenged by illustrations in the realm of identification, where photographs function well, they are more easily outperformed in most other communications realms. Ramachandran and Hirstein (1999) maintain that all art is evidence of the peak shift phenomenon as this is what the psyche responds to. However, this misses some important possibilities of visual art such as visual metaphor or synesthetic pictures which make visual – rather than recording (and distorting) the visual – those experiences received through the other senses. Illustrators have not missed these opportunities, as I will explain further below. First I want to depart from the narrow task of identification and voyage into the many broad tasks of categorization enabled by illustration.

**Categorization**

Details in a picture, and especially the exaggeration of salient details that distinguish specific examples from others, are effective in the communication task of identification. However, detail can become a distraction, which brings me to the second caveat signaled above: humans are not always trying to identify specific things, and do more with vision:

Depending on perceptual context and behavioural goals, objects are recognised in different ways, for example, as a cow, an animal, or living thing. The way objects are naturally recognised is by accessing information specific enough to differentiate them from similar objects (e.g., recognising an object as a cow rather than a horse or a buffalo) – a notion termed the basic or entry-level of representation [1,2]. However, part of understanding the meaning of an object also necessitates that more-general information is accessed – for example, the commonalities between similar objects that enable us to know that an object is part of a superordinate category (e.g., as an animal or living thing). (Clarke and Tyler 2015, p. 677)

For a great deal of any day people need merely to categorize at these basic and superordinate levels in their negotiations with the visual world (Rosch et al. 1976; Clarke
Note that, in these negotiations, “information specific enough to differentiate” is all that is required. The human brain is rarely concerned with all the detail the visible world supplies, not with the pictorial fidelity so effortlessly captured and furnished by photography, or exaggerated through caricature.

In these cases, where the visual communication task is to help the beholder connect the picture to basic and/or superordinate level mental concepts rather than to any specific level identity, in other words, to help categorize rather than identify, photography can be detrimental to the communication because its implicit specificity not only supplies too much information for efficient communication but also provides information in impertinent channels. For instance, a recent experiment concerned also with faces determined that photographs were outperformed by cartoons: “Participants identified emotions on briefly presented faces. Results showed that, at short presentation times, accuracy for identifying emotion on more ‘cartoonized’ images was enhanced. In addition, increasing contrast and decreasing featural complexity benefited accuracy” (Kendall et al. 2016). It should be noted that the task required was at the aforementioned basic level rather than the identification of individual persons; the task required participants to recognize emotions rather than identify individual persons. It may come as no surprise, then, that the benefits of distorting features through drawing have become a convention of particular illustration cultures. In manga and anime there exists a convention known as the “super-deformed” style. Entire cartoons may be made in this way, or visually realistic animations or comics may change in style, momentarily, to “super-deformed.” In these moments the drawings obscure a character’s specific identity in order to amplify, through a more general kind of peak shift effect, the emotion of the moment (Figure 1.1).

Reality is always furnishing a superabundance of detail. That is, more detail is available than the human visual system needs or can cope with. Visual noise must be jettisoned in order to get at the signals. In other words, the detail available in the visible world is not something to be taken on face value, but rather a difficult problem for

---

**Figure 1.1** “Hien” by Hien Pham (2018). An example of the super-deformed style which exaggerates moments of high emotion and foregrounds these ahead of the specific identity of a character. Source: © Hien Pham. Reproduced with the permission of the artist.
the human visual system to negotiate. To know what a thing is and where it is in a scene is not the same as to know who it is or which it is. As this is true of how humans interact with the real world, so it is true of how humans interact with visual representations of the world. Some means of illustration, applied in particular contexts, can reduce this abundance to more effective levels for the communication task at hand. A range of non-photographic picture-making methods appears to enable more effective visual communication because each has the ability to short-circuit various of the visual processes going on in the human visual system. For both tasks, categorization and identification, illustration offers modes that are more effective than visual information captured through the most high-fidelity means. Pictures communicate differently as a function of the degree of fidelity adopted by the illustrator, and this difference is key to the way that illustrations communicate. Distillation and exaggeration can actually communicate more powerfully to the psyche than “the real thing.” Several theorists have proposed that any images may be pictured with varying degrees of fidelity, and that a realism continuum may help in evaluating these for a range of purposes (Meggs 1992; McCloud 1993; Wileman 1993; Medley 2009). An example is given at Figure 1.2.

The Importance of Outlines

For most visual communication the outlines of objects are essential to perception and understanding of the elements of a visual scene. This outline is referred to in the scientific literature on vision as “contour.” In studies exploring the importance of contours in visual cognition, it is the outline or long contours in a visual scene or representation that are shown as most important for comprehension in most instances. In one experiment, Walther et al. (2011) found that simple line drawings featuring only long contours were as effective as color photographs in allowing beholders to decode visual scenes and distinguish objects (“beaches, city streets, forests, highways, mountains, and offices”) within these.

The same study found that short contours (interior detail within an object’s outline) were less helpful for effective recognition. Global (versus local) structure in the picture was more important. As gestalt theory has it, the whole is more important

![Realism Continuum](image)

**Figure 1.2** Medley (2013) (after Wileman, Meggs, and McCloud). An example of a realism continuum. As well as moving from concrete to abstract, the continuum could be used to evaluate the tasks of pictures from identification to categorization. Source: © Stuart Medley.
than the parts for effective cognition. This may be the case because in the wild, long
contours are more likely to define a boundary between figure and ground: that is, to
separate an object from its scene, or an object from other objects. Furthermore, in the
wild, the more distant from the viewer aspects of the terrain are, the more likely these
shorter contours will become invisible to the viewer (more can be found below on the
effects of distance with respect to visual realism). This general rule is referred to as
“texture gradient” and is a monocular depth cue. This ability to understand the
objects in a scene may work best for those kinds of scenes, or pictures of scenes, where
gross differences exist between the elements within the picture.

Of course, even in nature, depending on the position of the sun in the sky, objects
may be perceived only by their outlines. A person standing between the setting sun
and their beholder will display to the beholder only by the long contour around their
silhouetted shape, rather than through any of the shorter contours of the details inte-
rior to their outline. A silhouette may indicate then what kind of object is before the
beholder but not easily allow the beholder to solve what psychologists have called the
homogeneity problem, or which particular object, or who, is perceived.

Illustrators of course can accentuate these long contours and often do. Many modes
of illustration reduce visual detail and impose upon their subjects a drawn outline
where none exists in the real world. As well as being a resource-efficient means of
depiction, the intuition shown by many illustrators to approach depiction in this way
accords with the human visual system’s methods of processing natural images. As V.S.
Ramachandran writes: “when you look at a full-color picture, your attention is dis-
tracted by the clutter of texture and other details in the image. But a sketch of the
same object allows you to allocate all your attentional resources to the outline, where
the action is” (1999). In addition to a preference for long contours, the visual system
appears to be more alert to vertical and horizontal contours than it is to diagonal
contours. This sensitivity may reflect the natural distribution of objects’ orientations
(Geisler and Diehl 2002, p. 421; Girshick et al. 2011; Ganguli and Simoncelli 2014).
Any picture which plays to these bigger, hard-wired visual themes is more likely to
“score a hit” on the visual system. Pictures designed as distilled images might better
fit this evolved model than the images arriving from the real world, giving the eye and
brain, in effect, a higher-impact version of the visual world. The human visual system
seems able to cope with, even prefer, these pictures which distill visual realism into
something “less-real-than-real” (Medley 2010).

It is very difficult to determine where nurture takes over from nature; however,
at least some abstract visual cognition seems to occur prior to acculturation. Infants
presented with pictures of dots and a line in a face configuration give significantly
more attention to this picture than to a “non-face” configuration of the same
elements. This suggests that the first picture is relatively important to them and
perhaps even understood as a face (Fantz 1961; Morton and Johnson 1991; Crain
2015). The reduction of detail may be short-circuiting the visual processes: sending
a kind of “pre-processed” image to the brain. Children in particular seem more
prone to understanding emotion from simplified pictures. This may be because
children rely more on low-level or simplified features characteristic of cartoons to
process facial emotion (Gao et al. 2010), suggesting that the opposite might be
ture: that the reading of real faces comes less naturally than reading cartoon faces!
However, as I have said, for the beholder it matters what the task is: to identify the
person, or to categorize something about the face, for example, to understand what
emotion is being displayed.
The Less Real Image in the Wild

Once the eyes and brain evolved to be able to discern objects at a distance, stereopsis developed to allow better depth perception. There is an ideal viewing distance for human stereopsis (20/20 vision). The viewing experience, focusing at distances beyond 20 ft, or six meters, begins to verge on the “less-real-than-real.” At the very least, viewing conditions further than this distance might be regarded as “less-real-than-ideal.” Images seen at greater distances from this ideal effectively contain less detail than the ideal required for recognition. The photoreceptor mosaic that makes up the retina contains a limited number of receptors arrayed across a limited field. The resolution the retina allows limits the detail that can ever be supplied to the brain. Notwithstanding whether the beholder has 20/20 vision, the visual information retinas supply to the brain is reduced at increasing rates outside this distance. One does not merely keep perceiving, for example, a human face but with smaller details as the subject backs away. Instead the details of the face will begin to disappear. The whites of the eyes will not be apparent, then, further away still, the iris will vanish for the beholder. If the mouth is closed it will likely disappear, as will the nose and eyebrows, into the general color of the face. The image now available to the beholder is less than the ideal needed for identification, and if the subject is further away still, the beholder will not have the information available to even categorize the subject. In a sense this reality is less real; the image of the person at this distance is less “representational” than when they are in the room with their beholder. The possibilities regarding who or what the beholder is looking at are greater.

An illustration may capture the details, or lack thereof, in a way paralleling this growing distance. An illustration of a person, for example, lacking any detail interior to the long contour of the silhouette may still be recognized as a person, but not a particular person. In this regard it may be more fit for the purpose of communicating something about people in general rather than a particular person, age, gender, or ethnic group.

Perceptual Constancy

Complicating how distance and ambient light affect the images that the retinas present to the brain for recognition, orientation needs to be considered. As with changing distances, the beholder can never see exactly the same thing from precisely the same angle twice; the retinas are forever showing to the mind different pictures. That is to say, the mosaics that make up our retinas (Masland 2012) are registering shifting patterns of color from one moment to the next, even when the beholder is remaining relatively still. As Foster (2001) asks, “How, despite rarely experiencing the same image twice, do we recognize an object as being the same?” Some theorists make the case for novel stimuli in the eye being compared in the brain to the memory of something similar. Others argue that such a system could not deal adequately with radically novel stimuli, arguing instead for the importance of observing an object over time:

The environment as we experience it, is so structured that potentially very different images appearing in close temporal succession are likely to be views of the same object […] this temporal structure forms the basis of a tendency to associate images of objects together over short periods of time. (Wallis and Bulthoff 2000)
Similarly, Okamura et al. (2014) argue that an object viewed from different angles can be recognized and distinguished from similar distractors after the viewer has had experience watching it rotate. Interestingly they found that their subjects (monkeys) were able to recognize objects across rotations up to 60°, even though there had been no opportunity to learn the association between different views. In short, the mind may be relying on a range of faculties (other than simply a memory of individual objects) to explore its object hypotheses.

The nature of being in the world means the world’s infinite detail cannot be taken on face value. Details shift, as the human viewer shifts. Higher-order animals, including humans, need mechanisms to cope with the changing visual patterns which arrive upon, and are fed to their minds via, their retinas. These constant changes mean humans need a brain with a wide margin for error when it comes to assessing visual input. As Edelman (1999) says, “the matching of representations must be able to ignore differences in geometric details in favor of similarities among less specific features.” This margin for error may be so elastic as to allow extremely abstracted or distorted pictures to stand for more high-fidelity images or combinations thereof. As Kubovy observes (in Wagemans et al. 2012), the visual system must be able to cope with uncertainty, using a relaxed form of perceptual contour closure that can work reliably even for fragmented contours. If the brain is not solely relying on information from the retina to help a person make hypotheses and, ultimately, decisions, what is the brain relying on?

Psychologists speak of the presence of invariants as a clue to recognition. For instance, one could ask whether an object’s shape is orientation-dependent or orientation-invariant. Invariants are the aspects of the object or within the scene that maintain, for example, a configural relationship with each other regardless of variations in their orientation to or distance from the beholder. In terms of making pictures, invariants are those aspects of a scene under visual scrutiny that the human brain knows do not change even while the eyes are telling the brain “they’re changing”! It may be that the mental faculties which test for invariance are what allow a circle containing two dots and a line in a face configuration to represent “face” to the infant beholder. These mental faculties have been termed “perceptual constancies” (Ittelson 1951; Walsh and Kulikowski 1998), and have been further categorized into “size constancy,” “shape constancy,” “color constancy,” and so on. Shape constancy allows the brain to understand, for example, that the surface of a door is understood in the mind as a rectangle, regardless of whether the viewer is oriented perpendicularly to the door when it is shut or sees the edges of the door in some degree of converging perspective (as a trapezium) when the door is ajar. The difference between the observed and the understood is captured by artists when they depict these invariants. It is these aspects of pictures that allow simple “schema,” discussed below, to work as communicative graphics.

Perceptual constancies keep the brain from being endlessly surprised by the eyes. Because the beholder may be oriented toward the object at variable angles and distances, these constancies (and how they rely on invariance/invariants) mean that “near enough” in a representation is more than often good enough for object recognition, especially at the categorization level. Constancy behavior is defined as the attempt to maintain a world which deviates as little as possible from the world one has experienced in the past, which offers one the best chance of acting effectively (Ittelson 1951; Cassidy 2013). In this regard, the level of detail furnished by an illustration,
typically reduced in detail from a precise depiction of the real world, may not be a problem for the human visual system. In fact, an illustration may be a visual distillation of an object to its invariant components, to its essential schema, in effect averaging out a range of views of the thing. This range of views, captured as a kind of amalgam of possibilities, may better match the stored mental model of that thing.

**Gestalt**

As the above examples show, these constancies are mental faculties that appear to be directed at distilling the simplest, most consistent forms from a complex, and possibly changing, visual scene, in order to determine what belongs with what. As the psychologist Karl Popper established, the mind wants always to make the simplest hypothesis based on the available visual evidence, not because this simple hypothesis is most often right but because it is the easiest to change quickly as soon as further information becomes available. For recognition by the beholder, it’s not a matter of objects being illustrated to appear as they do in the real world, it’s a matter of them looking more like their real-world referents than they look like anything else. It is not difficult to conceive that an illustration can strip away (where appropriate) the visual detail and noise that in the wild inhibits this grouping process, and deliver the visual answer to the beholder. Grouping across a surface, and then between surfaces, allows a mind to know that, though the impression on the retinae is changing, the relationship between these elements is maintained. In this regard the principles are a test for invariance; three dots on a surface may change direction and orientation but keep moving together and maintain their relationships to each other. In this regard they are invariants that conform to the gestalt laws of common fate and proximity:

Historically, the visual phenomenon most closely associated with perceptual organization is grouping: the fact that observers perceive some elements of the visual field as “going together” more strongly than others […] Another is figure-ground organization. In general, grouping determines what the qualitative elements of perception are, and figure-ground determines the interpretation of those elements in terms of their shapes and relative locations in the layout of surfaces in the 3-D world. (Wagemans et al. 2012)

While there is no definitive list of gestalt laws, the key principles are configured around assembling conceptual wholes from perceptual parts:

When we look at the world, we usually perceive complex scenes composed of many groups of objects on some background, with the objects themselves consisting of parts, which may be composed of smaller parts, etc. […] Gestalt principles aim to formulate the regularities according to which the perceptual input is organized into unitary forms, also referred to as (sub)wholes, groups, groupings, or Gestalten (the plural form of Gestalt). (Todorovic 2008)

The beholder assembles parts of a scene in much the same way as the illustrator builds a picture using his or her “production scripts,” as will become apparent in the next section of this chapter.
Making the Invisible Visible

I have examined illustration and how it reflects human visual perception and cognition in roles suited to help the beholder to either identify or to categorize what’s depicted. Implicit in each case has been an assumption that illustration in some way represents the visible world. A broad survey of illustration will quickly reveal it need not. As I proposed in the introduction to this chapter, drawing and illustration can also embody experience received through other senses and make visible those aspects of human existence that are otherwise invisible. The assumption that illustration is about recording the visual, albeit in ways progressively more abstract, needs to be closely examined and problematized. De Sausmarez (2002) in Basic design: The dynamics of visual form, cautions against early training in high-fidelity drawing, in particular perspective, precisely because it gives primacy to the eye over other body-felt understanding of space:

Its value as part of foundational training is questionable, since by offering the student a ready-made recipe for achieving an illusion of space it curbs his spirit of enquiry and he may find himself drawing by academic theory of spatial projection and control rather than from sensation or personal observation. Not only does its logic and operation imply one fixed viewpoint, thereby freezing the entire visual field into static rather than dynamic relationships, but its “one-eyed” geometry rules out the sensation of physical involvement, bodyfelt interpreting of space and forms in space which has previously been described as an essential human concomitant of spatial experience and conception. (pp. 60–61)

Paradoxically, focusing on the invisible can open up an understanding of one of the major strengths of chirographic picture making. As Toni Johnson-Woods, author of Manga and President of the Pop Culture Association of Australia, says of drawing as communication:

The iconicity of images gives the illusion that all drawings are universal and easy to understand, since they can mimic the character of objects in our daily perception. Despite this, the ways “visual speakers” draw people remain just patterns in the minds of “artists.” (2014, p. 188)

One might also add, “and patterns in the minds of beholders” if the picture is to find a receptive audience. The perceptual idea of drawing makes sense because it accords with what Neil Cohn calls “the phenomenological experience of drawing” (2012, p. 168), but it does not account for key traits of drawing, in particular, why different people draw the same thing in different ways. Just as human experience arrives through any of the senses or through combinations thereof, so depiction can be derived from experiences received through senses other than sight. The illustrator may make visible these experiences rather than merely recording the already visible. There is not such a clear-cut distinction anyway between recording the visible world or making visible the images in the mind’s eye. To an extent we see what we expect to see given the environmental cues, that is, we impose our concepts from the mind outwards on to the scene we behold (Kok et al. 2012). Humans know about the world through all the senses and begin to understand it through image schemas upon which are built metaphorical concepts of how things in their world are interrelated.
As with the identity/category discussion above, my interest in schema and metaphor here will be focused through a psychology lens.

**Image Schemas**

How are these non-visible aspects of the world understood and developed into visual concepts in the first place? Pemberton and Nelson (1987) advanced the notion that “language development processes and art development processes are related in important ways despite surface differences.” Cohn (2012) adds that “drawing is similar to language in function, form, and development”; that the concepts that underpin the development of language are those which underpin developments in drawing. If this is so, one would expect to see culturally specific conventions in depictive approaches as well as variations between cultures. Daniel Richardson and colleagues from the Department of Psychology at Cornell University maintain that this is the case. They demonstrated strong agreement among participants asked to select or draw schematic representations, or image schemas (explained below), of concrete and abstract verbs. “For example, participants tended to ascribe a horizontal image schema to push, and a vertical image schema to respect” (Richardson et al. 2003).

With regard to variation between the representative schemas of different cultures, Wilson (1988) demonstrated that drawings within cultural groups may be based on a set of conventions common to the group and yet do not accurately reflect the actual appearance of those things in the real world. In this regard drawing can embody the communication of culturally specific schemas rather than being the reflection of visual perception. Notwithstanding some of the reasons why pictures and words are not two sides of the same coin, as I have argued elsewhere (Medley 2016), these visual schemas of artists may reflect a kind of visual parallel to a spoken dialect or accent.

**Image Schemas and Their Derivation**

Gibbs and Colston state that a large body of research supports the claim that image schemas are psychologically real (1995), yet, in spite of the visual implication in their naming, image schemas exist in the mind in a form more akin to what Steven Pinker refers to as “mentalese” (1997). Mark Johnson, one of the founding authors in 1987 of the concept of image schemas (though acknowledging the emergence of similar concepts in Immanuel Kant, Maurice Merleau-Ponty, William James, and John Dewey), explains their significance as:

> neither merely mental nor merely bodily, neither merely cognitive nor emotional, and neither thought alone nor feeling alone. All of these dimensions are inextricably tied up together in the perceptual and motor patterns of organism-environment interaction, which provide the basis for our patterns of understanding and thought. [...] What we call “mind” and “body” are not separate things. Rather, we use these terms to make sense of various aspects of the flow of our experience. Image schemas are some of the basic patterns of that flow. (2005, p. 18)

Image schemas are posited as the foundation of language, but Johnson sees image schemas as also pointing the way “to all forms of symbolic human interaction and expression” (p. 16). Lakoff and Johnson (1980) worked primarily in the area of linguistics; illustration theory or art history has not dealt adequately with or developed a parallel concept for image schemas. However, researchers are beginning to test the
value of these concepts in visual fields such as human–computer interaction (Bakker et al. 2012), music (Larson 2012), film (Ortiz 2014), and in interdisciplinary research spanning literature and comics (Forceville 2008; El Refaie 2015). A comprehensive illustration theory will need to have a position on these concepts as the building blocks for conceptualization and abstract thought upon which conceptual pictures are made.

Examples of these include the image schemas (capitalized here to reflect the convention of image schema theory), CONTAINER, COUNTERFORCE, SUPERIMPOSITION, SURFACE, BLOCKAGE, BALANCE, OBJECT, COMPULSION, SCALE, FULL, EMPTY, CONTACT, PATH, ATTRACTION, PROCESS, all of which emerge from perceptual experience provided by both touch (including proprioception and haptic perception) and vision (Popova 2005, p. 395). These image schemas are extremely abstract mental representations of the patterns of bodily interactions that build the way we understand the world. “The CONTAINER schema, for example, forms the basis of our daily experiences with houses, rooms, boxes, tea pots, cups, cars, etc.” (Hurtienne and Israel 2007, p. 130).

In the context of an illustration the CONTAINER schema might appropriate itself through the illustrator’s consideration of the canvas, page, or frame as a space in which to compose a picture, as well as in what the illustrator decides to depict. The size of the elements and their placement within that space will be in subconscious reference to the CONTAINER schema. Artists have a bodily relationship with the space in which they work, which derives from their physical place in the world: “every picture has a diagrammatic dimension, in and with it having a top and bottom, a center and periphery, close and distant relationships” (Troelsen 2017, p. 401). The space into which the illustrator places the illustration defines the illustration and separates it from what is not the illustration, and all of this before the illustrator even begins to depict the concept in mind, much of which will also rely on the illustrator’s conceptual metaphors. A brief case study is given below in relation to the infographic shown at Figure 1.4).

From Image Schemas to Visual Schemas

I have said that art history and illustration theory do not have a parallel concept but the more elaborate visual schemas which Gombrich referred to in Art and Illusion are built upon these more primitive, universal image schemas. For Gombrich there is a clear distinction between “seeing” things in the real world and “knowing” something about them, including knowledge of schema through which things can be depicted: “in all styles the artist has to rely on a vocabulary of forms and it is the knowledge of things that distinguishes the skilled from the unskilled artist” (2002, p. 246). Gombrich’s conception of schemas may hover between the artist’s subconscious and consciousness. His idea accords with Michael Kimmel’s (2005) conception of compound image schemas. Kimmel, a researcher in cognitive science based at the University of Vienna, explains that these compound image schemas result:

from stacking simpler image schemas on top of each other in a single imaginative locus through image-schematic superimposition. An example is superimposing a connective CONDUIT (i.e., a FORCE moving an ENTITY through a LINK) onto the space between two CONTAINERS to create the well-known folk-model of communication. (p. 303)
Gombrich says that an artist’s personal knowledge of these more complex schemas is the basis for that artist’s depictions and, as these schema are built up from personal experience with the world, the basis for their style. Kennedy, in “Metaphor and art,” adds that, “like Lego toys, objects are combinations of elements (surfaces), and the elements [of depiction] can be added or taken away ad infinitum” (2008, p. 454). Accordingly, a theory of illustration could, in addition to the ideas of representing the real world, look into the act of ordering existence that the illustrator undertakes when making a picture:

part of the revision of depiction lies in re-aligning pictures with patterns […] variations and orientations of a motif, serve to further principles of organisation in perception and cognition. It is not so much that depiction permits opportunities unavailable to observers of an object directly, but rather that we make more opportunities, and more of the opportunities, with depiction. (Bell 2001, pp. 26–27)

Metaphor

Because these schemas work across the senses, they connect vision to other domains of knowing and communicating about the world and our experiences in it. In these connections they form the basis for metaphorical thinking. As Hurttienne and Israel (2007) explain, “Although image schemas describe human experiences with the physical world their actual strength lies in their metaphorical extension for structuring abstract concepts” (p. 130). Lakoff and Johnson (1980) pointed out that “Metaphor is for most people a device of the poetic imagination and the rhetorical flourish [and] typically viewed as characteristic of language alone, a matter of words rather than thought or action. For this reason, most people think they can get along perfectly well without metaphor.” They demonstrated however that metaphor pervades everyday life, and that people’s understandings of the world and their place in it arise from the kind of bodies humans have and how these function in the environment: “Orientational metaphors give a concept a spatial orientation; for example, HAPPY is UP. The fact that the concept HAPPY, is oriented UP leads to English expressions like ‘I’m feeling up, today’. Such metaphorical orientations are not arbitrary. They have a basis in our physical and cultural experience.”

Metaphors, even at this basic, subconscious level as mere extensions of image schemas, contain a source domain and a target domain. Source domains are grounded in sensory-motor (body-felt) experience and are relatively concrete, while target domains are more abstract. In a phrase such as “I feel down,” the source domain is direction, a readily understood concept, and the target domain is the feeling of sadness, a notion that is as real but more difficult to articulate.

Conceptual metaphor theory is not without its controversies (Ortiz 2014; El Refaie 2015), which revolve around whether there are separate kinds of metaphor that need to be accounted for differently. Conceptual metaphors derive subconsciously and automatically. Even so, these may be the basis for the ability to think and communicate using more sophisticated metaphors such as those comprising the aforementioned “rhetorical flourishes,” where these also feature target and source
domains in an asymmetrical, concrete-abstract distribution. Certainly Richard Saul Wurman (1996) has said, in reference to sophisticated visual explanations in infographics, “you only understand something relative to something you already understand,” suggesting that the role of these designs is to make comparatively abstract concepts more concrete (an example of such an illustrated infographic is given at Figure 1.4).

Grady (1999) posited another kind of metaphor, which he called “resemblance metaphors,” as deriving from a physical or conceptual perception which is common in both domains, by the association of concepts with common features. Pictures may tend to invite visual comparisons more than concepts articulated through speech or writing. Because of the possibility of visual resemblance, illustrators will compare one image to another in order to see what so will afford. For example, the illustration for the record sleeve for the artist Michael Kentoff shown at Figure 1.3 replaces a headphone with an alarm clock because of the physical similarities between the two items. However, I suggest that these two kinds of metaphors may not be so clearly delineated: the context in which the resemblance metaphor is presented to the beholder may invite a search for a conceptual metaphor. Perhaps, if resemblance metaphors have any parallel in the verbal realm it is less with metaphor and more with rhyme or alliteration. These latter two have also been demonstrated as markers of language development (Missall and McConnell 2004; Liu et al. 2015). This will need to be addressed by a theory of illustration, especially since the problem may be compounded for illustration theory because different kinds of metaphor may coexist in the one work, as the example at Figure 1.4 demonstrates. As with image schemas being developed into the more elaborate visual schemas and production scripts of an illustrator,
so conceptual metaphors lead to an elaboration of metaphor in its better known guise of rhetorical flourishes.

To explore the idea of an illustrated metaphor to explain an abstraction, I return to the topic of “economic crisis” which I mentioned in the introduction to this chapter. Design team Always With Honor’s infographic, Greatest Bankruptcies in History, is an example not merely of recording the visible, but rather of making visible that which cannot be seen: these bankruptcies are invisible to the eye, nor can they be captured with a camera. Rather they have been made visible here through the illustrated embodiment of metaphors. These metaphors range from metaphors that may not have even registered as such during the design process, so commonplace is their application in human spatial cognition, to those clearly applied consciously by the illustrators.

The most obvious and clearly applied metaphor in the picture is the idea that bankrupt firms are akin to sinking ships. Bankruptcy, an abstraction, is the target domain in this metaphor, with sinking ships, a tangible, visible concept, as the source domain. In addition, the illustrators have clearly imposed a color logic on to the illustration where ships of like color belong to the same corporate sector. A key is provided to explain how the colors are applied. This in turn connects this novel infographic with a cartographic convention, pointing to another metaphorical link: infographics are like maps. Unlike a map, the horizontal spatial relationships between the ships do not
accord with a spatial relationship, scaled down, from the real world, but exist for reasons of legibility. Perhaps less obvious is that size of the ships accords with the size of the bankruptcies.

Another, more automatic metaphor is embodied in the downward direction necessary for the sinking metaphor. This is evidence of the image schema PATH, and the extended metaphor derived from this: down is negative. Extending the metaphor, the illustrators have run the timeline down the page to reinforce this downward trajectory. Less likely to have been a deliberately imposed strategy by the illustrators is the time/space metaphor applied on this y axis of the graph underpinning the infographic, since this is a key, but subconscious relational concept in human understanding of the world. The spatial understanding of time is a primary metaphor and the vertical spatial relationships of the ships are actually referring to a relationship in time. As Ortiz (2014) explains: “Primary metaphors are the minimal units of correlation-based metaphors; they are inherent in human nature and the result of the nature of our brain, our body and the world that we inhabit. We acquire them automatically and we cannot avoid them. Furthermore, as corporal experiences are universal, so are primary metaphors” (p. 5). As with all conceptual metaphors, the source domain (space) is more concrete than the target domain (time).

Edward Tufte might refer to this illustrative embellishment of financial-loss-as-sinking-ship as “chart junk” (1990). However, even a basic bar graph uses metaphorical translation. A bar graph still employs a rectangular shape (rather than a sea-going vessel) as a stand-in for a more abstract quantity such as financial worth. However, a flat rectangle lacks the sense of weight and size that illustrated ships imply, let alone the sense of being crewed by human beings, nor is a bar graph as novel and visually engaging as this more overtly metaphorical approach. This sensitivity to a body-felt understanding of a person’s place in the world is a core skill of an effective illustrator. As the illustrator eschews visual realism, a gestalt belonging between the elements in a picture may be made more apparent. Stepping away from visual realism allows the illustrator to make clearer these invariant elements within the visual depiction, as if clearing the dappled undergrowth of a forest to reveal the leopard’s spots. In fact, a kind of visual order may be imposed where none exists in the real world.

In more expressive illustration, such as the visual work for a children’s story, this imposition may be less intellectually ascribed to the elements in a visual scene: the imposition of commonalities such as uniform line-weight or limited color palette may arrive through an illustrator’s more intuitive or time-tested development of processes or selection of tools. Similarly, the approach to drawing people, for example the bug-eyed, overbite-prone characters of the Simpsons universe, may evolve more naturally through the practice of the illustrator, or indeed be a perceived “limitation” of the illustrator’s fine-motor abilities that have been nevertheless embraced and exploited to a particular effect.

In both of the above modes of illustration, which might broadly be labeled objective and subjective, visual realism has been left behind for something more effective. Schemas have remained intact and have become the focus ahead of faithful representation. In these modes order has been imposed on the visual display. This may be an order that unites, through logical application of color and shape, the disparate parts that share a relationship (other than a visual one) in a complex system. Or it may be an order that arrives through habitual use of tools and media, which shares more esthetics with fine art, but nevertheless holds the work together and supports a coherent, even if fictional, narrative.
Carroll (2001) says that a visual metaphor can be characterized in terms of three components: the material of the visual image, its subject matter, and its content, including its shaped properties: texture, color, line, mass, form, and the ways in which these are handled (p. 361). I would go further to state that the act of illustrating is itself metaphoric in the sense that I have been describing above. It makes visible that which is not. Its target domain is always something less tangible, either through spatial or temporal distance or some other abstraction, and its source is pencil on paper, or ink on board or paint on canvas. It is the act of making concrete.

Conclusion

Illustration exists in tension between perception and cognition, between seeing in and recording from the wild, but also making visible those feelings and understandings bubbling up from the deepest experiential and emotional recesses of a person. Furthermore, the very act of seeing in the real world is about abstracting the essence of scenes: working out what belongs in what superordinate category; and, if people are present, working out who is identifiable; and disregarding the other available visual information. In fact the human beholder is continually distilling their view to make sense of the environment and the important things within it. The illustrator can isolate, exaggerate, reduce, enlarge, outline, etc. these things on behalf of their works’ beholders.

In the same way that metaphor seems to the uninitiated to be about “rhetorical flourishes” but is actually grounded in the psychology of our being in the world, so too illustration, though it may sometimes appear to share esthetics with fine art, actually has deliberate tasks to accomplish. Its existence and its audience understanding depends on the psychology of vision. Even in surreal pictures which rely on coincidental resemblance (such as the picture at Figure 1.3 for example), the act of making such impossible things visual is a metaphoric act: it is making concrete on the page that which was merely thought. For the beholder the interest comes in allowing this artistic license and, to borrow another metaphor, going along for the ride.

Appendix: Character Design Workshop

The workshop described here is an application of the above research into teaching materials. The workshop is set up to question the value of life drawing and realistic proportions and propose instead steps intended to help participants focus on and accelerate toward their own illustrative style. At the conclusion of the workshop, participants will have two characters (for the potential for tension or conflict in a narrative) in a two-panel comic strip which demonstrates aspects of the characters’ traits. This workshop has been tested with beginners and experienced illustrators in the UK, Australia, China, the UAE, and Poland, with groups as small as six and as large as 40, to very positive responses. The workshop requires two hours.

In the first step participants are asked to draw a “normal” face for their age group. The participants may use a remembered formula for this, or they can move the elements around in the face (erasing and redrawing eyes, nose, mouth, face outline, etc.) until the face configuration begins to “disappear,” by which, it is explained, none of the individual elements draws attention to itself as too large or small or placed in the wrong location.
Given that style becomes apparent where depictive choices have been made, in the next step participants’ choices are made overt: they are urged to “break” the face drawing by moving the elements around within the configuration. Participants can begin by erasing the eyes and redrawing these elsewhere in the face, perhaps below the level of the mouth, or by moving the ears up to the top of the head. Participants learn that it is virtually impossible to make a non-face configuration as long as the elements are included; almost no amount of toying with the placement, even where the ears, now on top of the head, may suggest an animal more than a person, will lead to a face design that would not be useable as a character in a narrative. Paradoxically, the drawings have become more unique and memorable; more like useful character designs.

Participants then should take a schematic approach to the entire figure; drawing and reflecting on the “right way” for a human figure to look. They are asked, “Where do legs and arms bend?” and “are arms and legs thicker at the top or the bottom?” Again choices are prompted. Participants make multiple sketches, altering the relative drawn lengths of thigh and shin, upper and forearms, and inverting the thicknesses of limbs. These proportions appear to matter much less than other perceived physical properties. For example, having the fold lines around the elbow working in the right direction is much more germane to the “reading” of the drawing than whether the elbow is half way along the arm. Participants have discovered that where elements are placed in relation to each other is more important than actual proportion for readability.

Participants are asked to reflect upon at which points they felt that the tension between normality and uniqueness of expression felt “right” for them as picture makers. In this way participants begin to become aware of their own “production scripts,” that is, how they put figures together according to what they feel and know rather than what they see.

To develop their unique character designs participants are given a list of traits (e.g. lazy, eager, able, adventurous, rude, aloof) and told to choose three to five of these, or reflect on these to develop their own (within the tight time constraints of the workshop). The traits can contradict each other for complexity. Likewise, the visual aspects of the character design can either harmonize or contradict the chosen traits. Special emphasis is placed on the face and the eyes as these will be a focus for the audience. Participants are asked: “What if its hands are big? What if they’re small? What does this suggest? How does the physical attribute connect to the personality?”

Then they are asked: “Who or what will stand in your character’s way to create conflict and an interesting story?” Participants are reminded of the maxim that character is plot: what the characters want drives the narrative.

Escaping pictorial realism allows easier differentiation between characters. The more different each character is from each other, the easier the impression of consistency. The anthropomorphic character is common in comics for this reason. The danger is falling into stereotyping. The Japanese creator Miyazaki is adept at sidestepping this problem by creating his own animals, which become a sophisticated mix of the traits of others. Totoro is perhaps an owl, a cat, and a rabbit rolled together. He’s cuddly and wise.

**Caricature**

Caricaturing is another way to further develop the character designs that have evolved out of this process of “breaking drawings.” Participants are already seeing the outlandish results of deliberately “getting it wrong” with their schematic character
anatomy. Caricaturing their own creation will further exaggerate those differences that make the character unique. Participants compare their unique character design to a norm for that category of subjects, or, where no norm exists for the character created (for example, if it is a bizarre monster), a relatively simple composite may be derived through a combination of the common landmarks found in the two designs the participant has just made. Each of the two character designs can be compared to the composite and the differences from the composite should be exaggerated, for example, as in Figure 1.A1.

The reason for caricaturing the characters, rather than participants accepting their initial character designs, is to push further the possibilities of the design, especially those visual aspects that make it unique.

**Synesthesia**

The reliance on drawing schema and caricature rather than capturing the visible world as is has emphasized the role of illustration in making visible the internal world of the artist. The “body-felt” understanding of forms in space of de Sausmarez and the “expressive gestures of the experiencing body” Woodward alludes to are pointers to further means of fine-tuning the character designs. In the character design classroom activities participants are prompted to draw on their other senses to better refine their figures. Synesthesia can be accentuated in the shapes used to depict a character; angular corners on the character’s silhouette, for example, may suggest a more difficult or angry character. A smooth, curving outline may draw attention to a character’s relatively relaxed temperament, or it may help to make it appear more vulnerable to the fortunes of the narrative’s plot. Character poses can be synesthetic also. We know from body language when someone is sad or anxious. Participants are prompted to develop physical poses into a character’s design as permanent attributes.

Synesthesia can work on much more subtle levels too. The texture achieved through different coloring media and paper stocks and the line weight and character executed with the use of different pens or brushes may be used to harmonize with the traits of the character.
Finally, the two characters are placed into a two-panel comic strip, the most distilled comics form which allows the contrasting traits of the characters to direct the narrative. Comics are pinned up (or shown on an overhead projector) so that the group may see each other’s work.

Bibliography


Medley, S. (2016). Deliberately getting it wrong: accelerating towards finding a style for character design through the psychology of seeing. *4th International Conference on Illustration and Animation*. Edição IPCA.


